

Posudek oponenta diplomové práce

Název práce: **Rozpoznání slov diskrétního diktátu**

Autor práce: **Bc. Miloslav Kočí**

Předložená práce se zabývá dnes velmi aktuální problematikou automatizovaného rozpoznávání lidské řeči a porozumění obsahu počítačem. Předložený problém pokrývá velmi širokou a náročnou problematiku, je však do značné míry zjednodušen prvotním předpokladem rozpoznávání diskrétních slov. Zadaná problematika je v úvodní části práce analyzována a jsou zde v omezené míře teoreticky rozebrány vybrané metody používané v této oblasti. K této části práce mám nejzásadnější připomínky. Vzhledem k tomu, že předložená analýza se autorovi stala východiskem k návrhu vlastních postupů, zasloužila by si tato část podstatně hlubší teoretický rozbor. Za nedostatečnou považuji část zabývající se kvantizací signálu, možnostmi potlačení šumu obsaženého v signálu, či dopadem volby výběrového okna na další části řetězce zpracování. Velmi často se zde vyskytují krátké oblasti popisu (kap. 3. a 4.) bez vysvětlení další návaznosti a dopadu na výsledky rozpoznání. Přivítal bych také alternativní návrhy řešení a zdůvodnění výběru konkrétních metod. Koncepčně je ale úvodní kapitola členěna správně.

Za stěžejní považuji v kap. 4. pasáž věnující se rozdělení signálu do separátních rámců s předpokladem stacionárních parametrů a následné determinování příslušnosti rámců ke znělým, či neznělým hláskám. Není zde zcela jasné, zda kromě odečtu stř. hodnoty byla použita i celková normalizace signálu. V této části jsou bez bližšího popisu navrženy konstantní parametry počátku a délky trvání rámců a to bez vysvětlení důvodů a zejména diskuze dopadů na úspěšnost postupů popisovaných v kap. 5. a 6. Domnívám se, že zejména tato část se může značně podílet na dosažené nízké úspěšnosti celkového postupu.

Ve druhé části práce je krátce popsán systém záznamu a parametry záznamového prostředí a autor se následně věnuje popisu navržených algoritmů. Navržené algoritmy jsou vhodně prezentovány formou částí zdrojových kódů s komentáři a vysvětlujícím textem. V kapitole 8. je bez bližšího vysvětlení zmíněn lineární zesilovač. Není však uvedeno, zda se jedná o vlastní konstrukci, či převzatou a nejsou ani uvedeny naměřené výsledky nelinearity zesilovače ani jeho další parametry.

V poslední části jsou popsány navržené algoritmy a dosažené výsledky rozpoznávání na vzorku zaznamenaných slov, prezentované na celkem 8 řečnicích a dvou větách. Výstup algoritmu je prezentován pro tři odlišné způsoby omezení funkce přizpůsobení časové osy a jednotlivé výsledky jsou krátce diskutovány. Autor zvolil vhodnou podobu prezentace výsledků, není mi však zcela jasná volba „konstanty“ - viz. otázky.

Kapitola 7. je z hlediska práce celkem nadbytečná a zmíněný prostor mohl být využit ve prospěch detailnějšího rozboru problematiky v úvodu práce.

Celkově je práce napsána přehledným způsobem a autor volí správný sled navazujících postupů. Ve zdrojových kódech je vhodně dokumentován celý postup procesu rozpoznání, včetně podpurných skriptů usnadňujících opětovné spuštění. Práce se tak může stát východiskem pro další zájemce o tuto problematiku.

Otázky k obhajobě:

- V kap. 4.6 popisujete určování hranic slov. K jakému účelu je ve vaší práci tato informace dále využita?
- V kap. 2.2.1 uvádíte průměrné množství informace připadající na foném a z toho vyplývající datový tok. Dokážete tento údaj vysvětlit?
- V kap. 4.3 předpokládáte stacionaritu signálu v průběhu doby trvání rámce. Ověřoval jste na vzorku mluvčích, že tento údaj ve Vašem případě platí? Případně nebylo by možné velikost rámce přizpůsobit vzorku mluvčích (menší výpočetní náročnost/vyšší přesnost)?
- Jakým způsobem je určována „konstanta“ popisovaná v kap. 10 a jak se systém vyrovná se situací, kdy bude vysloveno slovo, které se nenachází ve slovníku (medián vzdáleností)?

Vzhledem k uvedeným připomínkám práci doporučuji k obhajobě a navrhuji ohodnocení klasifikačním stupněm VELMI DOBRĚ.

Pardubice 17.6.2010

Ing. Martin Dobrovolný